Ampersand 3 (2016) 13–20

Contents lists available at ScienceDirect

Ampersand

journal homepage: www.elsevier.com/locate/amper

"Kia ora. This is my earthquake story". Multiple applications of a sociolinguistic corpus

Lynn Clark*, Helen MacGougan, Jennifer Hay, Liam Walsh

School of Language, Social and Political Sciences, Department of Linguistics, University of Canterbury, Private Bag 4800, Christchurch 8140, New Zealand

HIGHLIGHTS

- This paper demonstrates how a corpus of spoken data can have multiple applications outside of linguistics.
- The QuakeBox corpus has been used in:
- The construction of a set of teaching resources for the high school curriculum.
- A study of the experiences and emotional responses of teachers.
- A project which seeks to examine water and waste activities in the wake of damaged sanitation infrastructure.

ARTICLE INFO

Article history:
Received 6 May 2015
Received in revised form
14 January 2016
Accepted 24 January 2016
Available online 4 February 2016

Keywords:
Applied sociolinguistics
Corpus
Monologues

ABSTRACT

This paper demonstrates how spoken data, collected using sociolinguistic methods, can have multiple applications outside of its original intended use within sociolinguistics. It can be a resource for tackling real-world problems, it can be a platform for community engagement and it can function as a source of data for academic research (both linguistic and non-linguistic research). The spoken data we describe is a new corpus of monologues called the UC QuakeBox corpus. First, we introduce and demonstrate the QuakeBox corpus, and outline some of the rewards and challenges associated with collecting stories in a manner that was purposefully and saliently in the public eye. Next, we focus on applications of the QuakeBox corpus by exploring case studies which are utilising data from the corpus for non-linguistic work. We situate this work within the wider field of applied sociolinguistics.

1. Introduction

The term 'applied sociolinguistics' was introduced to the linguistics community by Joshua Fishman (1970) and has come to be most commonly associated with establishing how research findings from sociolinguistics can be used by other fields, specifically with a view to tackling real-world problems (Trudgill, 1982:2). For instance, canonical work in applied sociolinguistics includes the sociolinguistics of second language acquisition (Schmidt, 1986), the social psychology of language (Giles, 1971a, 1971b; Giles and Powesland, 1975), language policy and planning (Haugen, 1966; Kloss, 1969; Fishman, 1974), discourse analysis (Labov and Fanshel, 1977; Tannen, 1982) and, increasingly, forensic linguistics (Nolan, 1983; for an overview of the connection between sociolinguistics and forensic linguistics, see Brunner, 2009). Each of these sub-disciplines of applied sociolinguistics itself now has a long and rich history.

More recently, another type of applied sociolinguistics has become popular under the umbrella of "outreach", "public engagement" or "impact". This has been inspired in part by the *principle of debt incurred* (Labov, 1982) and the *principle of linguistic gratuity* (Wolfram, 1993), but also, no doubt, by the recent emphasis placed on this type of activity by research funding bodies around the world.[1] This type of applied sociolinguistics mostly connects sociolinguistic data and research directly with the public, rather than with academics in other disciplines. For example, the North Carolina Language and Life Project (hereafter NCLLP)[2] has been

* Corresponding author.
E-mail address: lynn.clark@canterbury.ac.nz (L. Clark).

[1] For example, see the emphasis placed on public engagement by the Research Councils UK: http://www.rcuk.ac.uk/pe/.

[2] http://www.ncsu.edu/linguistics/ncllp/.

http://dx.doi.org/10.1016/j.amper.2016.01.001

collecting sociolinguistic recordings in North Carolina for more than two decades. The recordings have been the basis of valuable sociolinguistic work, but they have also been used for books and audio CDs written and constructed for the public (e.g., Wolfram et al., 2002), documentaries about dialectal diversity (e.g., Hutcheson, 2004; Rowe and Grimes, 2006), museum exhibits (e.g., Vaughn and Grimes, 2006) and in the production of school materials designed to raise awareness of dialect variation (Reaser and Wolfram, 2007).[3] An online archive, the Sociolinguistic Archive and Analysis Project (hereafter SLAAP),[4] was established as a web-based resource to store, catalogue and manage the increasingly large volume of recordings collected through the NCLLP (there are currently 1500 NCLLP interviews in SLAAP). Since then, other researchers have added their corpora to the website and it now houses over 4000 sociolinguistic interviews. Because of the web-based nature of this catalogue, it has a public presence. However, it was designed as a tool to aid sociolinguistic researchers, not for use by the general public. Access to the corpus is restricted and the access protocol on the website is clearly aimed at academic researchers: "Access to the SLAAP software and archive is password protected. Bona fide researchers can ask for and receive access to portions of the NCLLP's collection, dependent on the specific needs of the researcher and the human subjects permissions for the requested materials."[5] So while resources contained in SLAPP were used in the creation of the NCLLP's outreach materials, the catalogue itself is not intended for public exploration.

In the UK, the Diachronic Electronic Corpus of Tyneside English (hereafter DECTE)[6] is a similar web-based research platform to SLAPP which houses a large collection of sociolinguistic interviews from the Tyneside region. However, this project also has an accompanying public interface website called 'Talk of the Toon'[7] aimed at sharing a proportion of the DECTE recordings with the general public, with a specific target audience of those in education. Indeed, the Talk of the Toon website was designed with input from teachers and examiners in order to provide students and educators from primary to higher education with relevant materials (Corrigan, pc).

The examples of outreach work cited above from both the USA and the UK are mainly of sharing sociolinguistic interviews, collected for linguistic analyses, back with the community, and building training resources around these recordings. Indeed, this type of outreach work in which sociolinguistic interviews and dialect data are shared back with the community via the internet has become so popular that there is an edited book currently in preparation which describes the methods by which data have been created, digitized and exploited for similar outreach projects around the world (Corrigan & Mearns, forthcoming).

In this paper, we demonstrate how spoken data, collected using sociolinguistic methods, can have multiple applications. It can be a resource for tackling real-world problems (i.e. in the original use of the term 'applied sociolinguistics' described above); it can be a platform for community engagement (as in more recent examples of applied sociolinguistics or outreach) and it can function as a source of data for academic research (both linguistic and, increasingly, non-linguistic research). The spoken data we discuss is a new corpus of monologues called the UC QuakeBox corpus which is a collection of earthquake stories. In Section 2, we introduce and demonstrate the QuakeBox corpus, and outline some of the rewards and challenges associated with collecting stories in a manner that was purposefully and saliently in the public eye. In Section 3, we focus on applications of the QuakeBox corpus by exploring case studies which are utilising data from the corpus for non-linguistic work. Specifically, the QuakeBox has been used:

1. In the construction of a set of teaching resources for the high school curriculum which directly connects lessons across the Arts and Social Science curriculum to the devastating events which these pupils lived through and experienced first-hand (Clark and MacGougan, 2014)
2. In a study of the experiences and emotional responses of teachers, in their role as leaders and guardians in the wake of the earthquakes. One of the goals of this study is to explore opportunities for enhancing training and support mechanisms for teachers in high-stress environments (O'Toole and MacDonald, 2013)
3. In a project which seeks to examine water and waste activities in the wake of damaged sanitation infrastructure, and to explore the role of digital infrastructure in research activities (Butler, 2014).

The QuakeBox corpus has only recently been completed and released to the public, so the work discussed in this paper is primarily still ongoing.

2. Background to the UC QuakeBox corpus

2.1. The 2010–2011 canterbury earthquakes

A magnitude 7.1 earthquake struck the city of Christchurch and surrounding districts of North Canterbury, New Zealand, in the early hours of the 4th of September 2010. The city escaped without fatalities, though there was substantial damage to many buildings and infrastructure. Aftershocks continued to shake Christchurch and on the 22nd of February 2011, a hidden fault was jarred out of dormancy, resulting in a magnitude 6.3 earthquake that tore through the city at around lunchtime, causing 185 fatalities, some 7000 injuries, and the destruction of countless buildings, including much of Christchurch's city centre. Although the February earthquake was smaller in magnitude, it struck far closer to the urban area (only ~ 6 km from the city, compared with September's quake which was ~44 km from central Christchurch),[8] and it was shallower than September's seismic event had been. Also, ground acceleration readings measured more than twice the force of gravity – one of the highest such readings ever recorded.[9] The fault generated a great deal of vertical movement in addition to horizontal shaking, something few buildings (even those designed to be earthquake-resistant) are capable of withstanding. The result was the immediate destruction of many homes and buildings, including Christchurch's iconic cathedral, and extensive damage to a great many more, rendering much of the city's remaining infrastructure irreparable.

In the aftermath of these events, everyone who had experienced the quakes had a story to tell. These stories were diverse, and often dramatic, and people would tell their 'earthquake story' often. Several researchers at the University of Canterbury wanted to create a collection of these stories for three reasons. First, many

[3] For a more detailed summary of the NCLLP project and its outreach strategies, see Kendal & Wolfram (forthcoming).
[4] SLAAP: http://ncslaap.lib.ncsu.edu/index.php.
[5] http://ncslaap.lib.ncsu.edu/faq.php.
[6] http://research.ncl.ac.uk/decte/index.htm.
[7] Talk of the Toon: http://research.ncl.ac.uk/decte/toon/index.html.
[8] http://www.royalsociety.org.nz/media/Information-paperThe-Canterbury-Earthquakes-Scientific-answers-to-critical-questions3.pdf.
[9] http://www.gns.cri.nz/Home/News-and-Events/Media-Releases/Multiple-factors.

Fig. 1. The image on the left shows the exterior of the QuakeBox; the image on the right shows the interior. In the image on the right you can see that a door separates a sound-proofed, blue-felt recording booth from the observation room where research assistants monitored the recording in progress.

members of the public felt strongly that they wanted their stories to become a part of the public record, and be available for subsequent generations to learn from. There was a sense in which capturing, transcribing, and making these stories available to the public would be an important community service. Second, it was hoped that a collection of earthquake stories would provide a valuable repository for researchers across different disciplines interested in investigating the manifold personal and societal impacts of the earthquakes. Third, an archive containing multiple 'danger of death' monologues, each describing the same time and event, would be of particular value for sociolinguistic analysis.

2.2. The UC QuakeBox project

The UC QuakeBox project was formed as part of a collaborative endeavour between the New Zealand Institute of Language, Brain and Behaviour (hereafter NZILBB)[10] and the UC Canterbury Earthquake Digital Archive (hereafter CEISMIC).[11] The creation and structure of the corpus, including information about ethics and consent, is described in detail in Walsh et al. (2013). In an effort not to repeat Walsh et al. (2013), this section provides only a brief overview of the project and its goals.

The QuakeBox itself was a mobile recording studio built into a shipping container Fig. 1.

The QuakeBox was positioned at various locations in and around the city of Christchurch, and members of the public were invited to record stories of their experiences of the 2010–2011 Canterbury earthquakes. People came to speak freely and openly, for as long as they liked, about their experiences, both in the earthquakes and in dealing with the wide-ranging aftermath of these natural disasters. As an example, the link in (1) takes you to the public repository of Michelle Durham's story. She is a middle-aged female who recounts her and her husband's experiences during the February quake, and their efforts to re-build their community in the wake of the disaster:

(1) Michelle Durham's earthquake story

https://quakestudies.canterbury.ac.nz/store/part/79137.

Recording their own personal account so candidly and honestly was cathartic for many, and, given the sensitive nature of the stories, information was made available to enable participants to seek counselling, should they feel they required it. The stories were recorded in high quality audio and video, and they are mostly monologues - people were prompted with 'tell us your earthquake story' then left alone with the video camera to do just that. By the end of 2012 the QuakeBox project had recorded 722 stories in 13 languages.

Kendal (2011) explains that "the common practice in sociolinguistics is for individual (groups of) researchers to develop highly specialized, but closed, databases, which are not made widely available to outsiders" (2011:372). This is because sociolinguistic interviews sometimes capture sensitive information that the participants may not want to make public. Because the stories collected during this project were always intended to be shared publicly, it was possible to overcome this to some extent by requesting consent from participants for the many and varied ways in which their story could be made publicly available. Participants were allowed to choose from four research-related options, and five public-viewing options for sharing their story. They were also able to select which media they allowed to be accessed by whom (e.g. allowing researchers access to video, while restricting public access to audio-only; again, see Walsh et al. (2013) for further details). This perhaps sounds like a cumbersome consent form for participants to complete, but consistent with our impression that many people wanted to share their stories publicly, a total of 587 of the 722 stories were flagged by participants for full release i.e. they consented to have the audio, video and transcript released to the public and used in all ways specified. These 587 stories are available to view on the publicly-accessible UC CEISMIC Canterbury Earthquake Digital Archive website.[12]

A range of people with different social characteristics came to share their story. This can be seen from the age, gender and ethnicity information we have about speakers in the corpus (Tables 1 and 2).

The number of stories from people who identify as Maori ethnicity is low (only 3%) but the proportion of Maori residents in Christchurch is also low (only around 7%, significantly lower than some regions in New Zealand's north island) and it is unclear to what extent those who self-identified as mixed ethnicity or those choosing not to give ethnicity information were also of Maori descent. The QuakeBox corpus has also managed to attract a representative sample of both males and females from each age

[10] www.nzilbb.canterbury.ac.nz.
[11] www.ceismic.org.nz.
[12] https://quakestudies.canterbury.ac.nz/store/collection/235.

Table 1
Number of participants and their self-reported ethnicity in the public version of the QuakeBox corpus.

Ethnicity	Number of participants
NZ	431
NZ European	396
NZ Maori	19
NZ mixed ethnicity	16
Other	**117**
Declined to give ethnicity information	**39**
Total	**587**

'danger-of-death' stories (cf. Labov, 1972). Third, the corpus is different from most because the topic of the monologues is relatively uniform. This provides a degree of control over the topic of speech, something that is well-known to affect phonetic realization (Rickford and McNair-Knox, 1994; Gordon et al., 2004; Mendoza-Denton et al., 1999; Hay & Foulkes, forthcoming; Love and Walker, 2013). Finally, the data were collected in both high quality audio and video, unlike most current sociolinguistic corpora. These four factors are providing researchers at the University of Canterbury with the opportunity to explore within-speaker variation in

Table 2
Number of participants and their self-reported age and gender in the public version of the QuakeBox corpus.

Participant age	Female	Male	Declined to give gender information	Grand total
18–25	43	33		76
26–35	25	21		46
36–45	55	22		77
46–55	73	38		111
56–65	73	32		105
66–75	45	32		77
76–85	11	12		23
85+	3	6		9
Declined to give age information	6	17	40	63
Total	**334**	**213**	**40**	**587**

category, albeit with a slight over-representation of females in the middle age groups (see Table 2).

Finally, while there are many stories from the people of Christchurch, there are also stories from people who live elsewhere in New Zealand, or in other countries. Some people describe their earthquake experience from the perspective of someone who was not in Canterbury at the time, but who has since either returned to the city, or as someone who arrived as a visitor in the wake of the disaster. Only 44% of participants stated that they grew up in Christchurch or nearby districts in North Canterbury, and nearly 25% claim to have grown up outside of New Zealand.

Although not everyone agreed to make their story available to the public, they all agreed to allow their stories to be used by researchers at the University of Canterbury and so the corpus available for linguistic analysis contains 722 stories (approximately 120 h). In order to make these data available for linguistic analysis, the same practices of transcription and storage were adopted here as for the other corpora housed at the NZILBB (e.g. the ONZE database (Gordon et al., 2007), and the OLIVE database (Watson and Clark, in press)). The stories were first carefully transcribed and time-aligned in ELAN at the utterance level, then force-aligned with htk at the phoneme level and finally added to LaBB-CAT, a searchable online database developed and maintained by the NZILBB (for more information about these procedures, see Fromont and Hay, 2008, 2012).

There are several differences between the UC QuakeBox and more traditional types of sociolinguistic data which makes it an ideal resource for asking novel questions in linguistics. First, it is a collection of monologues.[13] This is useful because it opens up the possibility of exploring within-speaker variation in a way that is made much more difficult if we use a corpus of dyads i.e. the traditional sociolinguistic interview. Second, due to the nature of the topic, the speakers are unusually engaged in the monologues. It is, in some sense, the ideal sociolinguistic corpus – a collection of more controlled ways than is typically possible using traditional types of sociolinguistic data. For instance, the first author has been using the QuakeBox corpus to investigate individual variation and recency effects in phonological changes in New Zealand English (Clark and Walsh, 2014). Mountfort-Davies (2014) has questioned the extent to which group-level gender differences in the use of vocal creak are apparent across a range of individual speakers in the corpus. The multi-modal nature of the recordings is also allowing researchers to attempt to better understand the relationship between linguistic variation and gesture (cf Clark & Shelton, in prep; Gruber et al., in prep). This brief overview shows how this unique resource is being used to investigate novel research questions in theoretical linguistics and socio-linguistics. However, the content of the QuakeBox corpus is also interesting to researchers and practitioners working in other fields around earthquakes or natural disaster management.

3. Applications of the UC QuakeBox corpus

This section demonstrates several ways in which the QuakeBox data is being used by researchers and practitioners outside of linguistics. For this to happen there was a conscious effort made by the NZILBB and CEISMIC to 'spread the word' that this corpus existed, and that it was publicly available and free to use. For instance, there was considerable media coverage of the data collection phase[14] and a facebook page (https://www.facebook.com/UCQuakeBoxStories/) and twitter feed (@UCQuakeBox) were created which posted regular extracts from stories and links to the online video. Also, the QuakeBox corpus was hosted online by the UC CEISMIC Canterbury Earthquake Digital Archive project, which is a large project supported by a consortium of cultural and heritage organisations (including, for example, the National Library of New Zealand, the Ministry for Culture and Heritage and CERA – Canterbury

[13] We choose the term 'monologue' rather than 'narrative' to describe these stories to highlight the fact that they are single-talker recordings, not dialogues, and to downplay any expectation of recurring narrative structure.

[14] The data collection phase was recorded by TVNZ One News, Radio New Zealand (http://www.radionz.co.nz/national/programmes/afternoons/audio/2524140/southern-story-for-5-july-2012-the-quake-box) and The Press, Christchurch (http://www.nzilbb.canterbury.ac.nz/graphics/The%20Press_QuakeBox.pdf).

Earthquake Recovery Authority. For the full list of consortium partners, see here: http://www.ceismic.org.nz/consortium). Affiliating the corpus with this larger and more visible project undoubtedly helped to raise awareness of the QuakeBox outside of the University of Canterbury, and outside of linguistics.

Next, we outline three case studies showing how the QuakeBox database is being put to use in research and teaching outside of linguistics.

3.1. Sociolinguistic data as a resource for engaging high-school students

In the high school curriculum in New Zealand, particularly in English, there is a strong emphasis on encouraging teenagers to find and nurture a voice to tell their own stories and explore the stories of others.[15] The earthquakes that struck Canterbury in 2010-11 are among the most significant events in New Zealand's history. In collaboration with a local high school teacher from Christchurch, we have been exploring the possibility that connecting to these events in the classroom will encourage learners to take a more active role in learning because they will have been directly affected by these events themselves and so, of course, will all have their own earthquake stories to tell (Clark and MacGougan, 2014). We have developed the following 3 core modules[16] which teachers can use in order to get their students to attain certain Achievement Standards in the NZ curriculum:

Module description	Subject area
Developing a monologue as an oral text	English or Drama
Developing a monologue as a piece of creative writing	English
Developing a social action campaign	Social studies

An example of a teaching and learning unit that we have made available to teachers is called “Developing a monologue as an oral text”. This can be used with Achievement Standards for English at NCEA Year 11 or Year 12. Appendix 1 contains copies of the teaching pack that is available to students for this module; Appendix 2 contains the teachers' notes which help to guide the students through the module and makes sure that the students are achieving the key learning objectives necessary for completion of the module.

For the final assessment of this unit, students will write a script for a monologue of a character that survived the Christchurch earthquakes and perform this to the class. This teaching pack draws on examples of similar monologues from the QuakeBox corpus and guides students towards the point where they are able to construct their own monologue. A module such as this would take approximately 4 weeks to teach in high school. As a first step in creating the student resources (A1), we began by selecting a group of QuakeBox stories mainly from the 18–25 year old age group for students to read and listen to in more detail. In this activity, students identify and analyse language, gesture, and voice patterns from each of these stories and try to find recurring themes in danger-of- death or survival story monologues (for example, the stories are usually told in the first person, often in the past tense, and they often end with some moral lesson). This leads into an activity where students start to think more about what a dramatic monologue actually is and how it is performed. Again, the task is heavily focussed on exploring the language of monologues. Finally, students then start to think about how to apply this knowledge in order to construct and perform their own monologues, either by using their own experience of surviving the earthquakes or perhaps recreating the experiences of someone they know.

The teachers' notes (A2) are designed to make sure that the teachers are able to understand how each task that the students complete feeds into the overall assessment criteria for the module. It lists further resources where teachers can find more information on the language of monologues, or more information on the QuakeBox itself if they feel that they need some additional help preparing for teaching this module. Finally, the teachers' notes (A2) provide examples of what to look for in marking the unit.

In another example (not included in the appendices), we have developed resources for a teaching and learning unit called “developing a social action campaign”, this time for use in a social studies class room (again to be used with Achievement Standards for Social Studies at NCEA Year 11 or Year 12.) This module guides students through the process of developing a social action campaign that promotes a solution to an issue teens faced during the earthquakes or are facing in post-earthquake Christchurch. The underlying intention is to emphasize and promote teenagers' problem-solving skills. Students spend time looking for how teens identified changes that occurred in Christchurch as a result of the earthquakes and how the community responded to these changes, and they are encouraged to research different perspectives. The purpose of the campaign is to encourage teens to get involved and to put pressure on the government in the rebuild and resilience planning of Christchurch. Students are assessed on their ability to effectively develop and structure ideas, and use language features to command attention appropriate to the audience and purpose for writing (e.g. in the form of a newspaper article or perhaps a website).

3.2. Sociolinguistic data as a resource for understanding stress among teachers

Another example of work which has used the QuakeBox outside of linguistics is a study by O'Toole and MacDonald (2013) who are exploring the impact of the earthquakes on teachers, both emotionally and professionally. O'Toole and MacDonald (2013) use stories from teachers who took part in the QuakeBox project, and other stories that O'Toole herself collected, in order to better understand the stress teachers experienced in the wake of these natural disasters and their coping strategies. These researchers are particularly interested in how teachers dealt with their own emotions during the earthquakes. A common recurring theme discussed by teachers is how they felt that they had to regulate their own emotions and reactions to a life-threatening event in order to help the children. For example:

> (1) **Teacher**: “You just were on adrenalin. You just had to keep going and you couldn't um … .you didn't want to make the students frightened, so you couldn't look like you were frightened. That was the first thing – not to show fear, be frightened or cry” (O'Toole and MacDonald, 2013)

An interesting coping strategy which many of the teachers shared in their monologue was going into their ‘teaching bubble’ i.e. they report feeling emotionally well when they are teaching (their mood is better and energy levels are higher) so many of them threw themselves into their work as a way of coping with the aftermath of the events. Of course, this can lead to emotional

[15] http://seniorsecondary.tki.org.nz/English.

[16] These are in some ways similar to the resources developed in London for GCSE and GCE A level English Language students, using data collected by the Sociolinguistics Research Group at Queen Mary, University of London (http://linguistics.sllf.qmul.ac.uk/english-language-teaching).

fatigue, another theme explored in this work. Many teachers felt that they had a particularly difficult time after the quakes because they were doing far more emotional work with parents and students than before. Also, many of them didn't take a break from work (or they felt guilty if they did). Several schools were unsafe and so have closed or merged with other schools. This means that many teachers have a new workplace to adapt to and their own personal problems in the post-quake city have been pushed out of focus. For example:

(2) **Teacher**: "*I'm exhausted, I'm angry. I was up last night until midnight looking through EQC*[17] *documents and I'm arguing with them on top of my teaching role – about my home. … I'm trying to fix one thing in my life and that might mean (no longer teaching in) the school that I love and that I'm part of …. I'm sick of being in a broken situation*" (O'Toole and MacDonald, 2013)

This research is still in progress but it has the potential to contribute to methods of training teachers, both to maximise their capacity as leaders in high-stress situations, and also to minimise the amount of stress or "burnout" teachers might suffer. All of this also contributes to improving job satisfaction among teachers by furthering awareness of how the demands of their profession interact with their emotional state within the context of a traumatic event.

3.3. Sociolinguistic data as a resource for work in natural disaster management

Finally, one very real consequence of the Canterbury earthquakes for those who lived through the devastating events of 2010-11 was that many homes went without running water and adequate sanitation for months afterwards. "The Civil Defence and supporting agencies and authorities responded to the seriously damaged and non-functioning infrastructure by supplying residents with essential alternative water supplies. However, there is very little documented on how residents responded to the disruption of reticulated water supplies and the adequacy and use of alternative water sources in the immediate weeks after the earthquake" (Butler, 2014). A team of researchers from GNS science,[18] the University of Canterbury and Massey University[19] have been collaborating on a project which is mining datasets contained in the CEISMIC Digital Archive, including the QuakeBox corpus, to find discussions around water and waste use by households, individuals and communities in the immediate weeks after the Christchurch earthquakes. This project is particularly interested in using the UC QuakeBox data because its goal is to explore the role of digital infrastructure in disaster management research. Rather than generating new datasets, the task these researchers set themselves was to find and use existing data that had been generated by multiple and diverse sources in order to create knowledge and insights, in this case specifically geared toward waste and water use following a natural disaster. This is part of a larger project into 'Post-earthquake Functioning of Cities' (which is itself part of an even larger project - "Understanding Factors that Build Resilience in New Zealand") funded the Ministry of Business, Innovation and Employment in New Zealand.

For the research team, there were a range of datasets available for exploration but, according to Andrew Butler, the main research associate on the team, "The QuakeBox corpus proved to be a most useful resource for eliciting data on post-earthquake water use in the immediate weeks after the earthquakes. One particular strength was that we could 'ask questions' of the transcribed material" and so "drill down in detail to really understand the meaning of statements" (Butler, pc). Although these researchers were not linguists, the methods used were similar in nature to those used in corpus linguistics more generally. From an initial search term (e.g. 'water' or 'waste'), a number of additional key search terms (i.e. high frequency co-occurrences) were also derived. The UC CEISMIC analysts (who are experts in digital humanities, not linguistics) then assisted with writing what they have called 'scraper' scripts which lifted the key search string and a determined number of words either side. These are similar in nature to the practice of using concordance software to generate Key Word in Context (of KWIC) data. From here, the researchers interested in water management further explored this information for recurring patterns in discussions of water use. For example, Table 3 shows the recurrent mention of water being sourced from shops and supermarkets in the QuakeBox corpus. This allows the researchers to identify this as one of many strategies used by individuals following the earthquakes.

One particular direct question for this project was what lessons can be learned from the Canterbury earthquakes for other cities? The city of Wellington, New Zealand's capital city, is built on top of an active geological fault – the Wellington Fault. The research team had initially been keen to establish the volume of water used by Christchurch residents in the weeks and months following the earthquake in order to be able to make generalisations and possible preparations for future scenarios involving the similar-sized city of Wellington. It was not possible to retrieve this level of detailed information from the QuakeBox data but the team were able to provide thorough insights on water use and water sourcing techniques among the population which is undoubtedly invaluable information for disaster management planning.

The three case studies reported in Section 3 are necessarily brief because they are all examples of work in progress but they all show how a dataset that was collected primarily by and for sociolinguistics is being used in an entirely different way, with the potential to contribute to the development of new methodologies in different fields. Of course, we are keen to encourage more work in this vein. In an effort to make the QuakeBox corpus even more attractive to researchers and individuals outside of the University of Canterbury (and outside of linguistics), our next step is to further develop the public web interface of the QuakeBox corpus and make it searchable using LaBB-CAT (the full suite of LaBB-CAT search functions are currently restricted to those users who are granted access to the LaBB-CAT version of the corpus by the NZILBB). The functionality of the corpus will also be improved in other ways. For example, QuakeBox storytellers make frequent mention of places. We are working on ways to automate annotation of the geographical references, and link these to other resources – both internally to other QuakeBox stories featuring the same places, and externally to relevant CEISMIC resources. For example, a QuakeBox user researching an interview that mentions the Christchurch Cathedral will be able to select a link that leads to images, video, and newspaper pages related to the cathedral, held in CEISMIC. Conversely, a CEISMIC user researching the Cathedral will be able to select a link that takes them to the relevant part of a transcript in the QuakeBox, which they can click to listen or view. This should make the QuakeBox corpus more valuable to the public and to researchers across a range of disciplines.

[17] EQC = The Earthquake Commission; Government organization in New Zealand which provides natural disaster insurance for residential property).

[18] GNS science website: http://www.gns.cri.nz/.

[19] Led by David Johnston (http://www.massey.ac.nz/massey/expertise/profile.cfm?stref=428930).

Table 3
Extracts from the QuakeBox corpus showing how people talked about sourcing water from supermarkets/shops (Butler, 2014).

Speaker name	Extract
AP518_FoxSwindells.eaf	um. so we thought well w ~ we'll go for a walk to the supermarket and see if it's open and grab some **water**.
AP518_FoxSwindells.eaf	so we. decided to go for a walk to the supermarket and get some **water**. cos we. were. we never had an emergency kit we still don't.
NB177.eaf	**water** food so we stopped at the. local. supermarket. and bought containers of **water** and heaps of baked beans and spaghetti and. packet stuff and toilet rolls.
AP518_FoxSwindells.eaf	um and I think we did in the end but they'd l ~ by d ~ that stage they had limits on the **water** amount you could **buy** when we first went they didn't have any limits – um -
QB750_Gilly.eaf	my friend from Christchurch she had her credit card so - we racked up a whole lot of **water** and. [tuts] milk and bread and. loaded the car up -
UC212YW_CarolinStechel_.eaf	I guess they wanted to. um start reopening as soon as possible because people were. you know looking for **water** and all sorts of things because -
WF2607_Kurt.eaf	got up in the morning and d ~ we had no power or **water** so. a~ I jumped in the car and went down to Pak n Save to. to get some water -
SU2058LJW_Annie.eaf	and ahh her husband had gone down to get the paper and get some **water** and supplies.

4. Conclusion

This paper has outlined some of the linguistic and non-linguistic applications emerging from work on a new corpus of stories, the UC QuakeBox corpus, which was collected and transcribed using sociolinguistic methods. We have shown that these data are being used by researchers and educators outside of linguistics in several interesting ways. Specifically:

(1) In line with recent examples of applied sociolinguistics which share sociolinguistic interviews back with the community (cf. the papers in Corrigan and Mears, forthcoming), the QuakeBox corpus is being used in the creation of teaching resources for high school students (Clark and MacGougan, 2014)
(2) The QuakeBox corpus is also being used as a resource in disaster management research (O'Toole and MacDonald, 2013; Butler, 2014).

This second application is perhaps the most interesting as this is a rather unexpected and, we think, unique way for sociolinguistic data to be used. In some ways, this is similar to the original sense of the term 'applied sociolinguistics' because the work described here is using sociolinguistic data to tackle real-world problems, but it is also different because these data are not necessarily being used to tackle language-related problems (such as language planning or language teaching).

Kendal (2011:372) explains that "since sociolinguistic datasets have typically been developed in order to research a specific question or set of questions, it has often been assumed that once the original questions have been studied in depth there is not further interest in the datasets themselves". This paper has shown that there is indeed interest in further exploring the data that we often take for granted in sociolinguistics, such as the canonical 'danger of death' stories that linguists have been collecting since the 1970s (Labov, 1972)), and this data may well be of interest to researchers in other disciplines. Of course, in some ways, the UC QuakeBox corpus is a unique dataset because the participants are all describing the same event. However, we believe that the rich array of topics discussed in sociolinguistic interviews more generally may well be of interest to researchers in other disciplines. With only a few small changes to current sociolinguistic data collection protocol (such as expanding the participant consent form to allow at least parts of the data to be made available to others), we believe that more sociolinguistic corpora might find a voice in wider domains of use.

Acknowledgements

The QuakeBox project is a collaboration between UC CEISMIC and NZILBB at the University of Canterbury, both of whom provided considerable support for the project. Special thanks to Derek Bent, Liz Grant, Jeanette King, Paul Millar, Viktoria Papp & Kevin Watson and, of course, to all of the other NZILBB and UC CEISMIC staff and students who contributed in various ways. The project benefitted greatly from the support of Canterbury Community Trust, PBT Transport, Leighs Construction, Tourism New Zealand (for the donation of the QuakeBox itself) and the University of Canterbury Summer Scholarship scheme for funding several student interviewers. We would also like to thank Scott Lloyd, Robert Fromont and Emma Parnell for their technical and administrative support. The project also benefitted from the advice or services of Jayne Austin, Gary Busch, Rob Morris, Jessica Petersen and Rob Stowell. Finally the QuakeBox corpus project would not have been possible without the generous support of the host sites, at: Eastgate Mall, New Brighton Library, Brooklands, Lyttelton, Sumner, the Canterbury A&P show, Westfield Riccarton, and Cashel Mall (Re:Start).

Thanks also go to the audience at NWAV 43 who provided valuable feedback on this paper during a special session entitled "Novel impact pathways in variationist linguistics: Learning for and from specific societal contexts". The session was organised by Heike Pichler & Ashley Hesson.

Finally, thanks to the Royal Society of New Zealand's Marsden Fund (grant number UOC1309) for funding some of this work.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.amper.2016.01.001.

References

Brunner, Elizabeth Gentry, 2009. The study of variation from two perspectives. Lang. Linguistics Compass 3 (3), 734–750.
Butler, Andrew, 2014. Initial Report from Existing Data Sources on Christchurch Residents Water Use in the Weeks after the February 22nd Earthquake (submitted for publication).
Clark, Lynn and Dominic Shelton (in prep) "Circumscribing the sociolinguistic variable – should we include gestures?".
Clark, Lynn, MacGougan, Helen, 2014. Resources for students and teachers from the QuakeBox Canterbury earthquake stories. In: Paper Presented at the New Zealand Linguistics Society Conference. Waikato University (November).
Clark, Lynn, Walsh, Liam, 2014. Recency effects on word-medial/t/in New Zealand English: initial observations. In: Paper Presented at the 15th International Australasian Speech Science & Technology Conference (SST). New Zealand,

Christchurch (December).
Corrigan, Karen and Adam Mearns (eds.) (forthcoming). *Creating and Digitizing Language Corpora, Volume 3: Databases for Public Engagement*. Basingstoke, UK: Palgrave Macmillan.
Fishman, Joshua A., 1970. Sociolinguistics: a Brief Introduction. Newbury House Publishers, Rowley, Massachusetts.
Fishman, JoshuaA. (Ed.), 1974. Advances in Language Planning. Mouton, The Hague.
Fromont, Robert, Hay, Jennifer, 2008. ONZE Miner: the development of a browser-based research tool. Corpora 3 (2), 173–193.
Fromont, Robert, Hay, Jennifer, 2012. LaBB-cat: an annotation store. In: Proceedings of Australasian Language Technology Association Workshop, pp. 113–117. http://www.aclweb.org/anthology/U12-1015.
Giles, Howard, 1971a. Ethnocentrism and the evaluation of accented speech. Br. J. Soc. Clin. Psychol. 10, 187–188.
Giles, Howard, 1971b. Patterns of evaluation to R.P., South Welsh and Somerset accented speech. Br. J. Soc. Clin. Psychol. 10, 280–281.
Giles, Howard, Powesland, Peter F., 1975. Speech Style and Social Evaluation. Academic Press, San Francisco, NC.
Gordon, Elizabeth, Campbell, Lyle, Hay, Jennifer, Maclagan, Margaret, Sudbury, Andrea, Trudgill, Peter, 2004. New Zealand English: its Origins and Evolutions. Cambridge University Press, Cambridge.
Gordon, Elizabeth, Maclagan, Margaret, Hay, Jennifer, 2007. The ONZE corpus. In: Beal, Joan, Corrigan, Karen, Moisl, Herman (Eds.), Creating and Digitizing Language Corpora. Volume 2: Diachronic Databases. Palgrave Macmillan, Basingstoke, pp. 82–104.
Gruber, James, Jennifer Hay, Jeanette King and Lucy Johnston (in prep) "Gesture-prosody synchrony in the QuakeBox videos".
Hay, Jennifer & Paul Foulkes (accepted, forthcoming) "The evolution of medial (-t-) in real and remembered time". To appear in *Language*.
Haugen, Einer, 1966. Linguistics and language planning. In: Bright, W. (Ed.), Sociolinguistics. Mouton, The Hague, pp. 50–71.
(producer.) Hutcheson, Neal, 2004. Mountain Talk. North Carolina Language and Life Project, Raleigh, NC.
Kendal, Tyler and Walt Wolfram (forthcoming) "Engagement Through Data Management and Preservation: The North Carolina Language and Life Project and the Sociolinguistic Archive and Analysis Project". In Karen Corrigan and Adam Mearns (eds.) (forthcoming), *Creating and Digitizing Language Corpora, Volume 3: Databases for Public Engagement*. Basingstoke, UK: Palgrave Macmillan.
Kendal, Tyler, 2011. Corpora from a sociolinguistic perspective. Rev. Bras. Linguística Apl. 11 (2), 361–389.
Kloss, Heinz, 1969. Research Possibilities on Group Bilingualism. A report. Center for Research on Bilingualism, Quebec.
Labov, William, 1972. The isolation of contextual styles. In: Sociolinguistic Patterns. University of Pennsylvania Press, Philadelphia, pp. 70–109.
Labov, Fanshel, William & David, 1977. Therapeutic Discourse: psychotherapy as conversation. Academic Press, New York.
Labov, William, 1982. Objectivity and commitment in linguistic science. Lang. Soc. 11, 165–201.
Love, Walker, Jessica & Abby, 2013. Football versus football: effect of topic on/r/ realization in American and English sports fans. Lang. Speech 56 (4), 443–460.
Mendoza-Denton, Norma, Hay, Jennifer, Jannedy, Stefanie, 1999. Probabilistic sociolinguistics: beyond variable rules. In: Bod, Rens, Hay, Jennifer, Jannedy, Stefanie (Eds.), Probabilistic Linguistics. MIT Press, Cambridge MA, pp. 97–138.
Mountfort-Davies, Morganna, 2014. A Gender-based Account Creaky Voice Phonation in New Zealand English. Unpublished honours research project. University of Canterbury.
Nolan, Francis, 1983. The Phonetic Bases of Speaker Recognition. Cambridge University Press, Cambridge.
O'Toole, Veronica, MacDonald, Alexandra, 2013. Emotional impact of the earthquakes on teachers: 'unsung heroes'. In: Paper Presented at UC CEISMIC Mini-conference. University of Canterbury (June).
Reaser, Jeffrey, Wolfram, Walt, 2007. Voices of North Carolina: Language and Life from the Atlantic to the Appalachians. North Carolina Language and Life Project, Raleigh, NC.
Rickford, John, McNair-Knox, Faye, 1994. Addressee- and topic-influenced style shift: a quantitative sociolinguistic study." In: Biber, Douglas, Finegan, Edward (Eds.), Perspectives on Register: Situating Register Variation within Sociolinguistics. Oxford University Press, Oxford, pp. 235–276.
Rowe, Ryan, Grimes, Drew, 2006. This Side of the River (producers.). North Carolina Language and Life Project, Raleigh, NC.
Schmidt, Richard W., 1986. Applied sociolinguistics: the case of Arabic as a second language. Anthropol. Linguist. 28 (1), 55–72.
Tannen, Deobrah (Ed.), 1982. Analyzing Discourse: Text and Talk. Georgetown University Press, Washington.
Trudgill, Peter, 1982. Introduction. In: Trudgill (Ed.), Applied Sociolinguistics. Academic Press Ltd, London.
Vaughn, Charlotte, Grimes, Drew, 2006. Freedom's voice: celebrating the Black Experience on the Outer Banks. Museum Exhibit. Outer Banks History Center & North Carolina Language and Life Project, Manteo, NC.
Walsh, Liam, Jen, Hay, Derek, Bent, Liz, Grant, Jeanette, King, Paul, Millar, Viktoria, Papp, Kevin, Watson, 2013. The UC QuakeBox Project: creation of a community-focused research archive. N. Z. Engl. J. 27, 20–32.
Watson, Kevin and Lynn Clark (in press, to appear 2015) "The Origins of Liverpool English". In Raymond Hickey (ed.) Listening to the Past. Cambridge: Cambridge University Press.
Wolfram, Walt, Dannenberg, Clare, Knick, Stanley, Oxendine, Linda, 2002. Fine in the World: Lumbee Language in Time and Place. North Carolina State University, Raleigh, NC.
Wolfram, Walt, 1993. Ethical considerations in language awareness programs. Issues Appl. Linguistics 4, 225–255.